Moving on from Foliage Height Diversity: determining maximum entropy in 3-dimensional variables

R. Valbuena[1], S. Adnan[2], M. Maltamo[2], L. Mehtätalo[3], R.N.L. Ammaturo[4], T. Lovejoy[5]

[1] Bangor University. School of Natural Sciences, UK
Email: r.valbuena@bangor.ac.uk

[2] University of Eastern Finland, School of Forest Sciences, Finland
Email: adnan@uef.fi; matti.maltamo@uef.fi

[3] University of Eastern Finland, School of Computing, Finland
Email: lauri.mehtatalo@uef.fi

[4] University of Strathclyde, Mathematics and Statistics, UK
Email: nla32@cam.ac.uk

[5] UN Foundation and George Mason University, Virginia, USA
Email: tlovejoy@unfoundation.org

1. Introduction

McArthur and McArthur's (1961) foliage height diversity (FHD) is widely used for determining structural complexity, from LiDAR vertical height (H) profiles (Lefsky et al. 2002, Vierling et al. 2008, Simonson et al. 2014). FHD has however largely failed to disentangle the relationships between the ecosystem structural diversity and biodiversity, with early reports such as those from Thomas Lovejoy (1972) in the Amazon not finding evidences in the light of FHD. It remains unclear whether FHD is the most suitable means to determine the structural complexity of ecosystems.

The calculation of FHD involves layering the vertical profile, which is essentially unnatural to describe a continuous variable (X) such as height, and involve subjective steps such as the determination of the size of these layers, from which the value of FHD obtained is ultimately dependent upon. This is because FHD is based on Shannon's (1948) entropy index, which was not originally designed to describe continuous variables, but meant for abundance data for categorical variables. In Adnan et al. (2021) we provided a mathematical framework for determining maximum entropy in 3D remote sensing datasets based on Lorenz curves and Gini (1921) coefficients (GC) determined from theoretical continuous distributions, intended to replace FHD as entropy measure in vertical profiles of LiDAR heights. This framework was developed for 1-dimensional variables (1D; X) such as tree heights, and 2-dimensional variables (2D; $Z \propto X^2$) such as basal areas, and hereby we extend it to 3-dimensional variables (3D, $Z \propto X^3$) such as volumes.

Structural complexity is an essential morphological trait of forest ecosystems, complementary to others like vegetation height or cover (Schneider et al. 2017, Fahey et al. 2019, Valbuena et al. 2020). But the means to measure the structural complexity of forests lacks consensus (Neumann and Starlinger 2001, Lexerød and Eid 2006, Valbuena et al. 2012). Two types of approaches, those measuring entropy (McArthur and McArthur 1961) versus those measuring variability (Weiner 1990), have effectively been merged in the framework presented in Adnan et al. (2021), by by showing how maximum entropy can be flagged up from values of a variability measure such as the Gini coefficient. Formal deductive proofs for maximum entropy at $GC = 0.33$ for 1-dimensional variables (Adnan et al. 2021), and $GC = 0.50$ for 2-dimensional variables (Valbuena et al. 2012, 2107), have been presented, which hereby are extended toward the value of $GC = 0.60$ for 3-dimensional variables.

2. Methods

Let $\mathrm{E}[X]$ be the expectation a random variable X with probability density function (p.d.f.) $f_X(x)$, cumulative distribution function (c.d.f.) $F_X(x)$, quantile function (inverse of the c.d.f.) $F_X^{-1}(p)$. The Lorenz curve $L_X(p)$ specifies the accumulated proportion of the total of X that is attributed to a given accumulated share of the population ordered by increasing X:

$$L_X(p) \;= \frac{\int_0^p F^{-1}(t)dt}{\mathrm{E}[X]}, \text{for } 0 \leq p \leq 1 \qquad (1)$$

Published in: Markus Hollaus, Norbert Pfeifer (Eds.): Proceedings of the SilviLaser Conference 2021, Vienna, Austria, 28–30 September 2021.
Technische Universität Wien, 2021. DOI: 10.34726/wim.1861
This paper was peer-reviewed. DOI of this paper: 10.34726/wim.2021

The Gini coefficient is the twice area between the Lorenz curve and the diagonal line $L_X(p) = p$, which is thus assessed with the integral:

$$GC_X = 1 - 2\int_0^1 L_X(p)\,dp \qquad (2)$$

When considering the distribution LiDAR heights $X = H$, the Lorenz curve $L_H(p)$ specifies the proportion of total accumulated ranked heights (usually in decreasing order, but it can be either). If considering 2-dimensional variables, such as basal area $X = BA = D^2$, then it gives the proportion of basal area for ranked trees (best in increasing order, to express competitive dominance, following Valbuena et al. 2013). We can also be interested in 3-dimensional variables, such as volume $X = V = HD^2$. The methods consist in mathematical proofs demonstrating values of Lorenz curves (1) Gini Coefficient (2) that can be used to characterize maximum entropy from theoretical distributions of 3-dimensional variables, which can be employed to substitute the use of FHD and avoid its unnatural partitioning of continuous variables into layers.

3. Results

3.1 Maximum Entropy in 3-dimensional variables: volume

Tree volumes are also calculated from a transformation of other dimensions $V = aHD^2$. Again, given the scale-invariability property of Lorenz curves, and thus we can consider the Lorenz curve and Gini coefficient of transformation $Z = X^3$ when $X \sim U(0, \theta)$.

The c.d.f. and p.d.f of the transformed variable are:

$$F_{X^3}(z;\theta) = \begin{cases} 0, \text{ for } z \le 0 \\ \sqrt[3]{z}/_{\theta}, \text{ for } 0 \le z \le \theta^3 \\ 1, \text{ for } z \ge \theta^3 \end{cases} \qquad (3)$$

$$f_{X^3}(z;\theta) = \begin{cases} \frac{1}{3\theta\sqrt[3]{z}}, \text{ for } 0 \le z \le \theta^3 \\ 0, \text{ otherwise} \end{cases} \qquad (4)$$

Thus, the quantile function and expected value of Z are:

$$F_{X^3}^{-1}(p) = \theta^3 p^3 \qquad (5)$$

$$E[X^3] = \frac{\theta^3}{4} \qquad (6)$$

Substituting these in Equation (1), the Lorenz curve becomes (Figure 1):

$$L_{X^2}(p) = \frac{\int_0^p \theta^3 t^3\,dt}{\theta^3/4} = \frac{\theta^3 p^4/4}{\theta^3/4} = p^4 \qquad (7)$$

And thus, substituting in Equation (2), the Gini coefficient of a uniform distribution becomes:

$$GC = 1 - 2\int_0^1 p^4\,dp = 1 - \frac{2}{5} = \frac{3}{5} \qquad (8)$$

Hence, for any variable $Z \propto X^3$ that is proportional to the third power of X, such as of V, the $GC_{X^3} = 0.60$ corresponds to the maximum entropy of X.

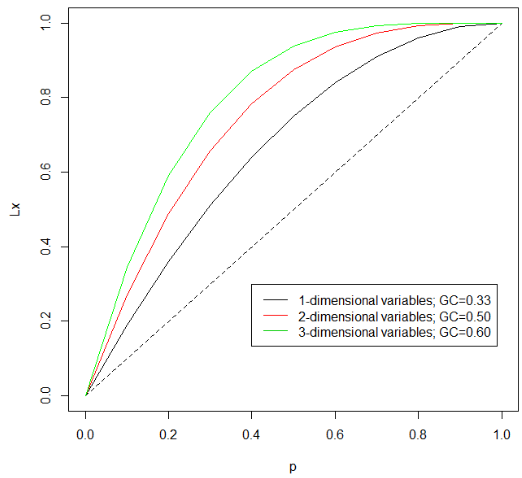

Figure 1: Lorenz curves for 1, 2 and 3-dimensional variables.

4. Discussion

In previous contributions we have showed a threshold of interest which flags up maximum entropy in forest ecosystems at the Gini Coefficient value of $GC_{X^2} = 0.50$ (Valbuena et al. 2012, 2017). In Adnan et al. (2021) we further deducted that the value $GC_H = 0.33$ can be used when interested in the study of LiDAR height profiles. In this contribution we show how higher order extensions can be further deducted, and show the formal proof for the maximum entropy value of $GC_{X^3} = 0.60$ applicable to 3-dimensional variables. In order to achieve these generalized conclusions, we use theoretical distribution functions and show how their parameters propagate into Lorenz curves and values of the Gini Coefficient directly dependent on those parameters. Further extensions can be similarly deducted based on ecological assumptions on ecosystem distributions.

These threshold allows to compare the entropy of the ecosystem using a statistic of dispersion, arguing that for continuous variables it is more correct to use the Gini Coefficient because it avoids the factitious binning step required when computing FHD (McArthur and McArthur, 1961). Gini coefficient is less computationally demanding than FHD, but in Valbuena et al. (2012) we also showed that it is conceptually better.

References

Adnan S, Maltamo M, Packalen P, Mehtätalo L, Ammaturo R.N.L., and Valbuena R, 2021, Determining maximum entropy in 3D remote sensing height distributions and using it to improve aboveground biomass modelling via stratification. *Remote Sensing of Environment*, 260: 112464.

Fahey RT, Atkins JW, Gough CM, Hardiman BS, Nave LE, Tallant JM, Nadehoffer KJ, Vogel C, Scheuermann CM, Stuart-Haëntjens E, and Haber LT, 2019, Defining a spectrum of integrative trait-based vegetation canopy structural types. *Ecology letters*, 22 (12): 2049–2059.

Lefsky MA, Cohen WB, Parker GG, and Harding DJ, 2002, Lidar remote sensing for ecosystem studies. *BioScience*, 52 (1): 19–30.

Lexerød NL, and Eid T, 2006, An evaluation of different diameter diversity indices based on criteria related to forest management planning. *Forest Ecology and Management*, 222: 17–28.

Lovejoy TE, 1972, Bird species diversity and composition in Amazonian rain forests. *American Zoologist*, 12: 711-2.

McArthur RH, and McArthur JW, 1961, On bird species diversity. *Ecology*, 42: 594–598.

Neumann M, and Starlinger F, 2001, The significance of different indices for stand structure and diversity in forests. *Forest ecology and Management*, 145(1-2): 91–106.

Schneider FD, Morsdorf F, Schmid B, Petchey OL, Hueni A, Schimel DS, and Schaepman ME, 2017, Mapping functional diversity from remotely sensed morphological and physiological forest traits. *Nature communications*, 8 (1): 1–12.

Shannon CE, 1948, A mathematical theory of communication. *The Bell System Technical Journal*, 27: 379–423 and 623–656.

Simonson WD, Allen HD, and Coomes DA, 2014, Applications of airborne lidar for the assessment of animal species diversity. *Methods in Ecology and Evolution*, 5 (8): 719–729.

Sung PY, and Bera AK, 2009, Maximum entropy autoregressive conditional heteroskedasticity model. *Journal of Econometrics*. 150 (2): 219–230.

Valbuena R, Packalén P, Martín S, and Maltamo M, 2012, Diversity and equitability ordering profiles applied to study forest structure. *Forest Ecology and Management*, 276:185–195.

Valbuena R, Packalen P, Mehtätalo L, García-Abril A, and Maltamo M, 2013, Characterizing forest structural types and shelterwood dynamics from Lorenz-based indicators predicted by airborne laser scanning. *Canadian journal of forest research*, 43 (11): 1063–1074.

Valbuena R, Maltamo M, Mehtätalo L, and Packalen P, 2017, Key structural features of boreal forests may be detected directly using L-moments from airborne lidar data. *Remote Sensing of Environment*, 194: 437–446.

Valbuena R, O'Connor B, Zellweger F, Simonson W, Vihervaara P, Maltamo M, Silva CA, Almeida DRA, Danks F, Morsdorf F, and Chirici G, 2020, Standardizing Ecosystem Morphological Traits from 3D Information Sources. *Trends in Ecology & Evolution* 35 (8): 656–667.

Vierling KT, Vierling LA, Gould WA, Martinuzzi S, and Clawges RM, 2008, Lidar: shedding new light on habitat characterization and modeling. *Frontiers in Ecology and the Environment*, 6 (2): 90–98.

Weiner J, 1990, Asymmetric competition in plant populations. *Trends in Ecology & Evolution*, 5 (11): 360–364,